

Film Market-Entry Analysis

So what's in it for you?

For incoming movie investors and producers:

- Benefits and risks analysis of the movie market
- Briefing on each determining factor
- Recommendations for revenue growth and movie success



Revenue Focused



Data Driven



Market Conscious

Data Source

- We got our data from Kaggle "The Movie Dataset"
- The Full MovieLens Dataset consisting of 26 million ratings and 750,000 tag applications from 270,000 users on all the 45,000 movies

Tables:

- **movies_metadata.csv**: The main Movies Metadata file. Contains information on 45,000 movies featured in the Full MovieLens dataset. Features include posters, backdrops, budget, revenue, release dates, languages, production countries and companies.
- **keywords.csv**: Contains the movie plot keywords for our MovieLens movies.
- **credits.csv**: Consists of Cast and Crew Information for all our movies.
- **links.csv**: The file that contains the TMDB and IMDB IDs of all the movies featured in the Full MovieLens dataset.
- **links_small.csv**: Contains the TMDB and IMDB IDs of a small subset of 9,000 movies of the Full Dataset.
- **ratings_small.csv**: The subset of 100,000 ratings from 700 users on 9,000 movies.

Overall Analysis Structure

Data Exploration

Movie Market Overview
and General Data
Analysis

Specific Attribute Analysis

Analysis on How Particular
Attributes Contribute To A
Movie's Final Revenue

Model and Final Summary

Data-Driven Predictions and
Business Insights

Core Questions to Consider...

**What are the determining factors for movies to gain high revenue/
market success?**

**For a set certain budget, what is the recommended action plan in
terms of detailed attributes such as selection of actors, producers and
keyword taglines?**

01

Data Exploration

Movie Market Overview and
General Data Analysis

Data Overview & Exploration

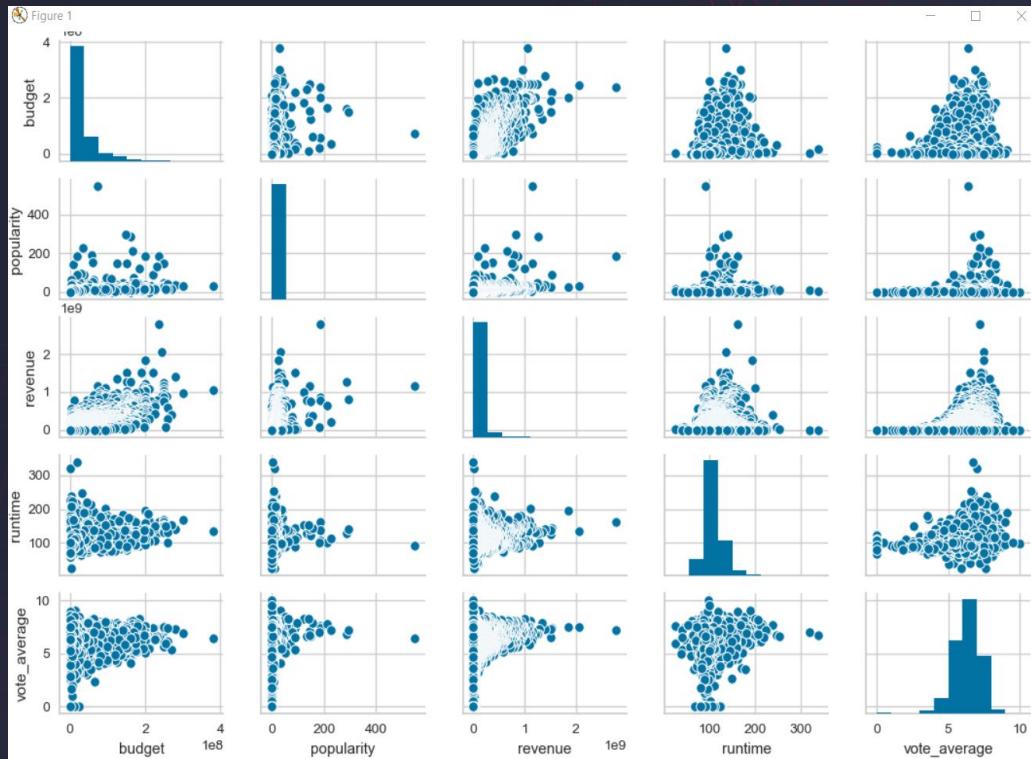
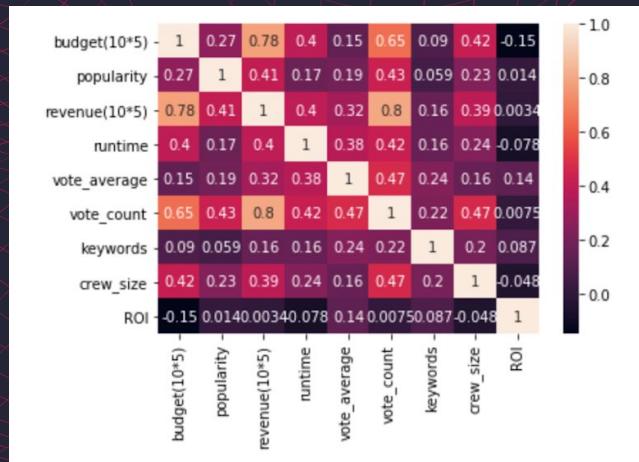
Out[123]:

	budget(10^5)	popularity	revenue(10^5)	runtime	vote_average	vote_count	keywords	crew_size	ROI
count	1191.000000	1191.000000	1191.000000	1191.000000	1191.000000	1191.000000	1191.000000	1183.000000	1191.000000
mean	481.697249	14.001272	2008.920236	106.988245	6.202183	1389.654912	9.449202	35.67202	7.199575
std	570.214076	23.375952	2690.385939	20.194679	0.855184	1839.799948	6.182428	36.63335	23.065576
min	1.140000	0.293373	0.029700	0.000000	2.800000	2.000000	1.000000	1.00000	-0.999406
25%	90.000000	7.628446	334.611305	93.000000	5.700000	238.000000	5.000000	11.00000	1.151412
50%	250.000000	10.491734	971.386860	103.000000	6.200000	630.000000	8.000000	20.00000	2.706469
75%	660.000000	14.523663	2613.836015	118.000000	6.800000	1751.500000	13.000000	46.00000	5.640529
max	3800.000000	547.488298	27879.650870	320.000000	8.500000	12269.000000	37.000000	182.00000	419.522723

On average the movies are able to ge a positive ROI with the minimum in negative 0.99 and maximum in positive 419.52. The maximum budget is 3.810^8 while the maximum revenue is 2.7810^9 . Both budget,popularity and revenue contains outliers

Correlation chart

Revenue has a strong correlation with budget and vote_count.



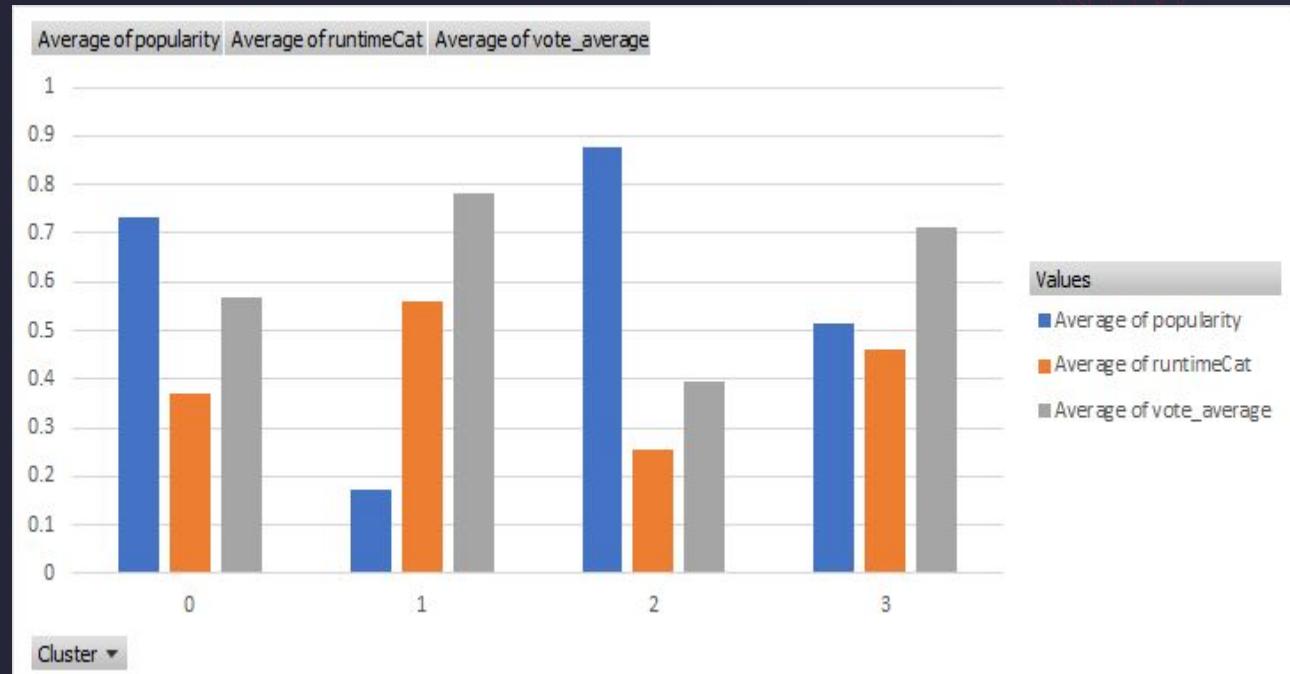
Average revenue of each cluster (KNN Clustering)

Cluster 2: 161.63 million

Cluster 0: 54.08 million

Cluster 3: 20.12 million

Cluster 1: 11.47 million



02

Specific Attribute Analysis

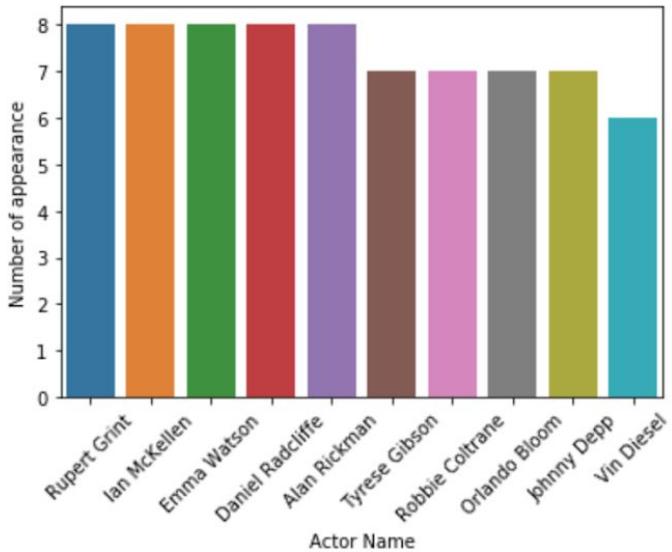
How Each Factor Contribute To A
Movie's Final Revenue

ACTORS

How the selection of actors contributes to a movie's revenue

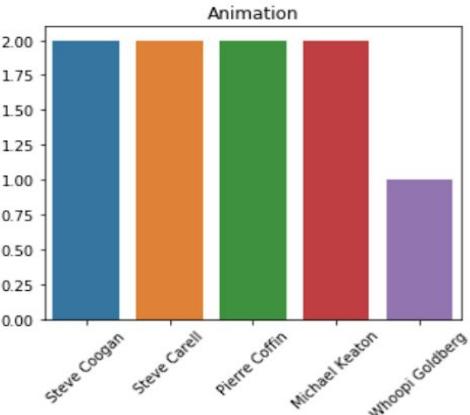
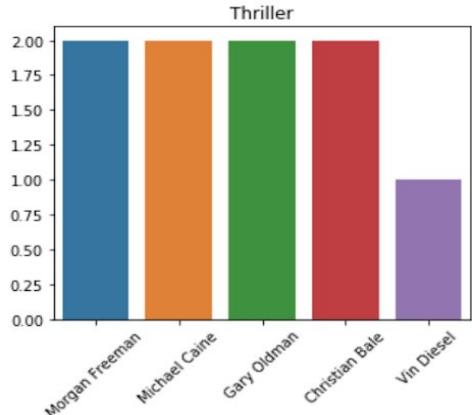
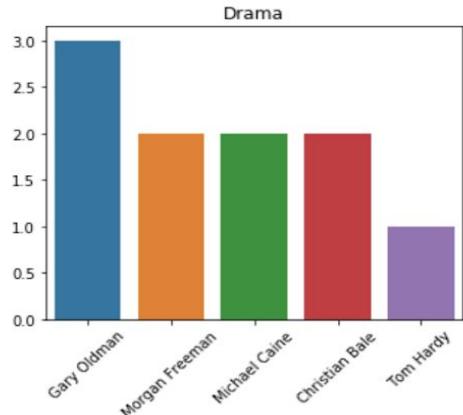


Top 10 Actors Based on Appearance Frequency in top 100 movies since year 2000



Top 5 Actors in Top 3 Genres by Revenue

```
In [600]: print_graph(df_r)
```

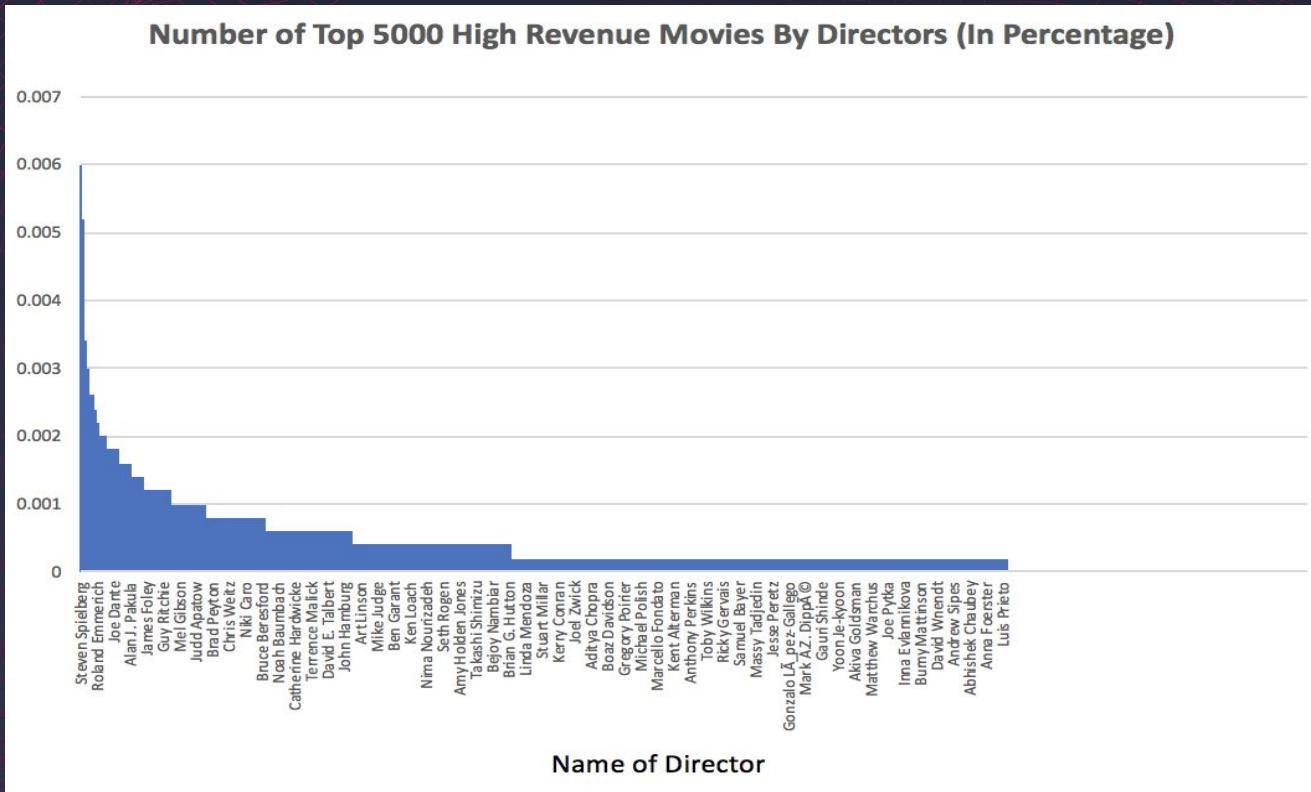


DIRECTORS

How the selection of directors
contributes to a movie's revenue

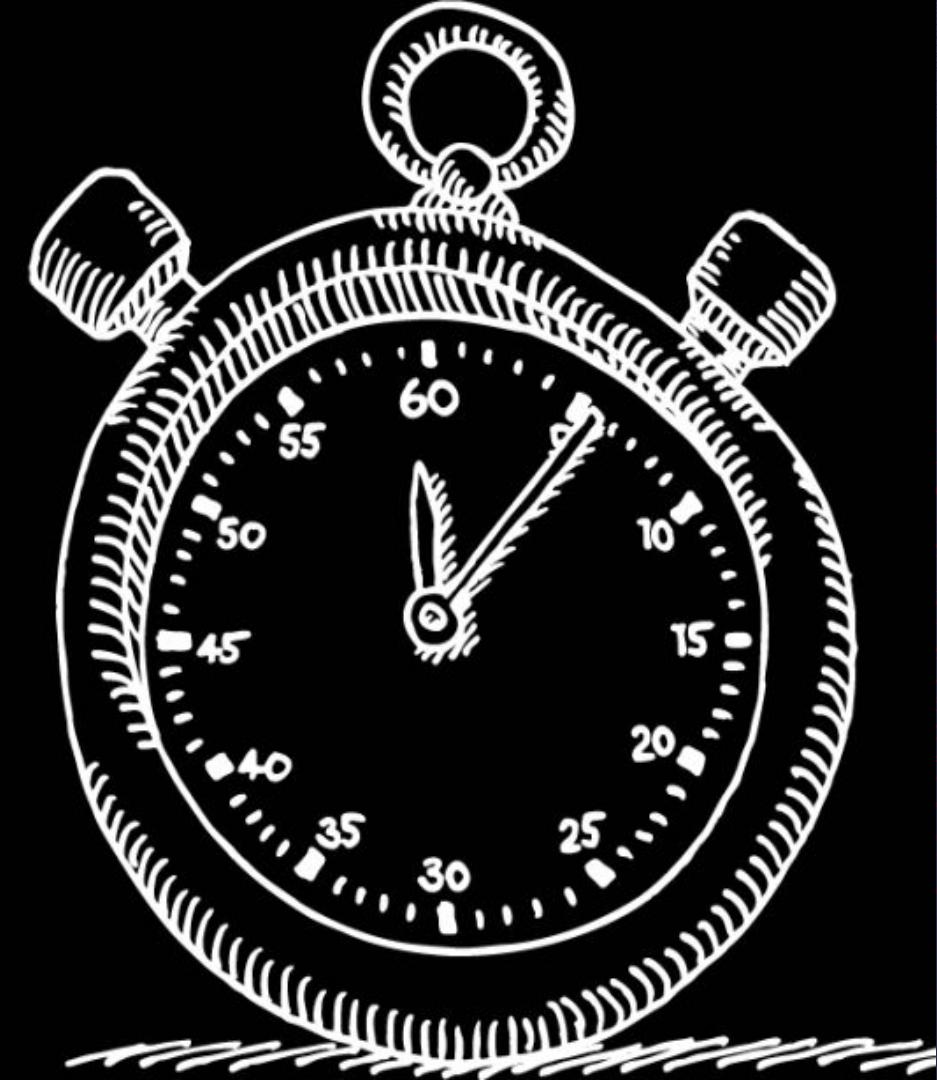


Top Directors Associated with High Revenue Movies



TIMING

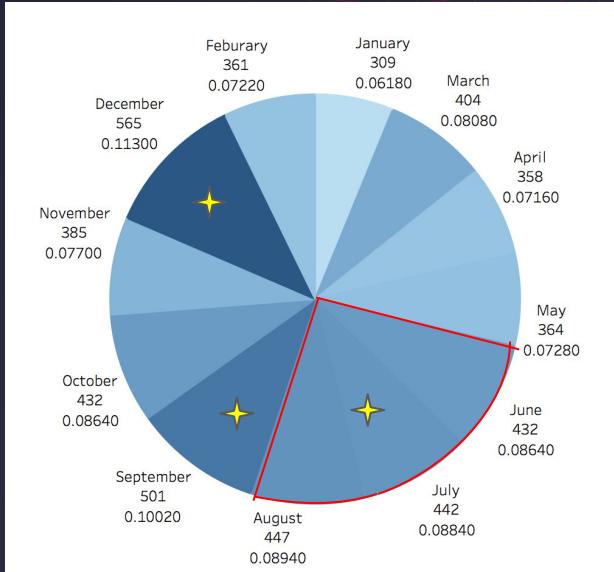
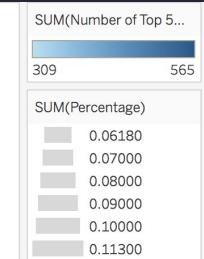
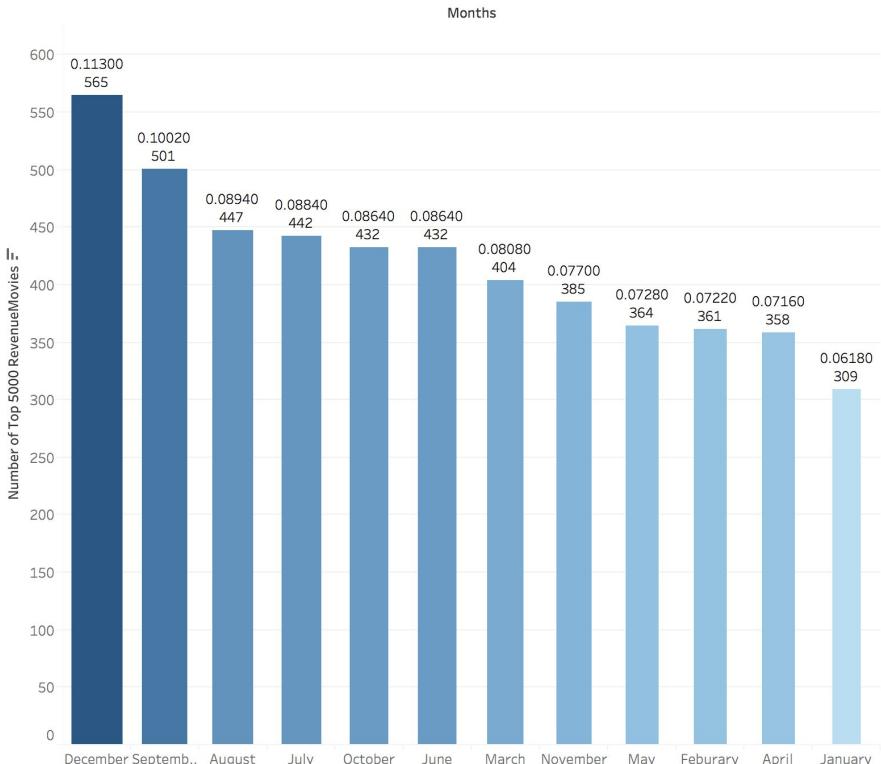
How the release timing decision contributes to a movie's revenue



Best Timing to Release and Premiere Movies

(Release Time Distribution for High Revenue Movies)

Best Timing to Release and Premiere Movies

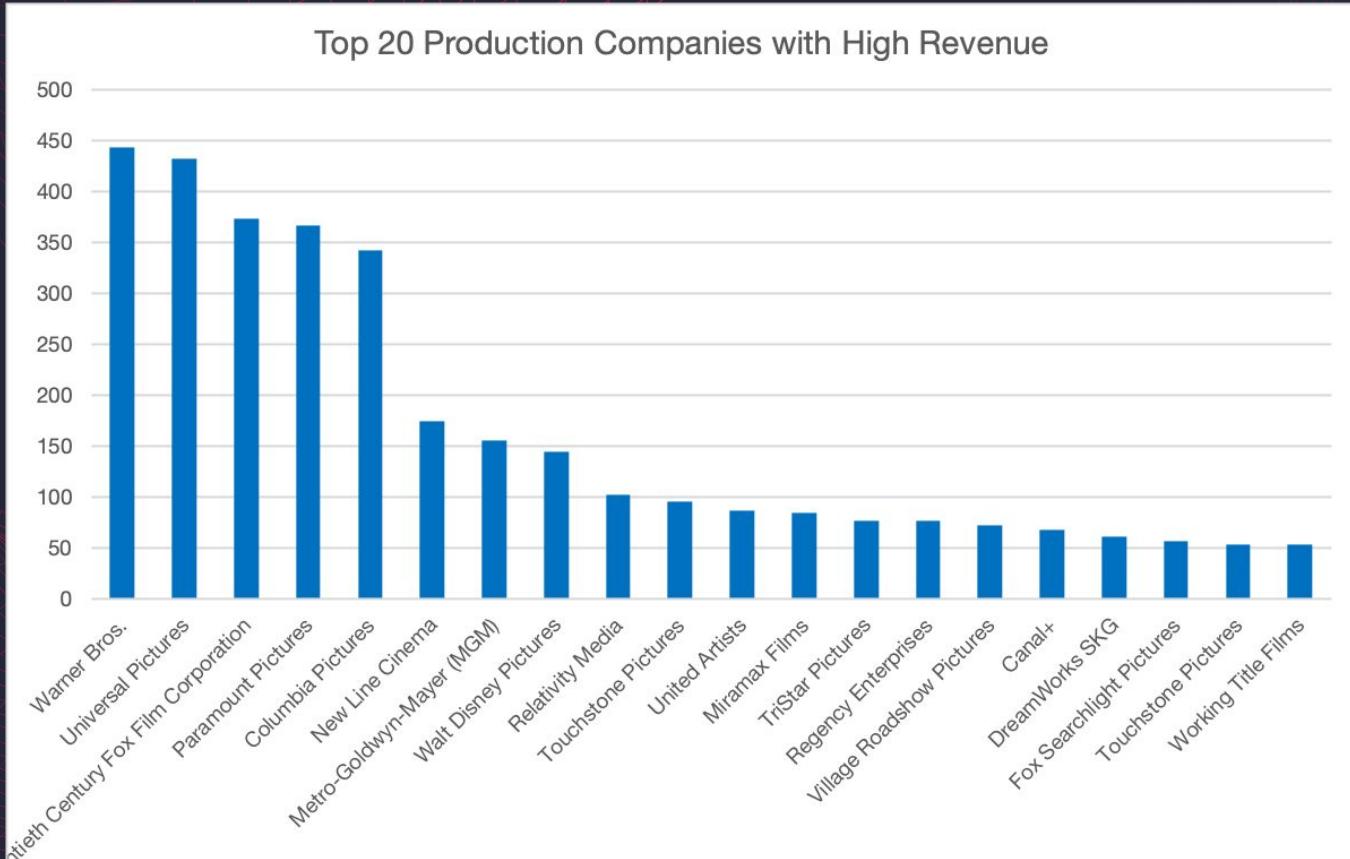


PRODUCTION COMPANY

How the selection of production company contributes to a movie's revenue



Top 20 Movie Production Companies with High Revenue



TARGET COUNTRY MARKET AND LANGUAGE

How the selection of target country market and language contributes to a movie's revenue



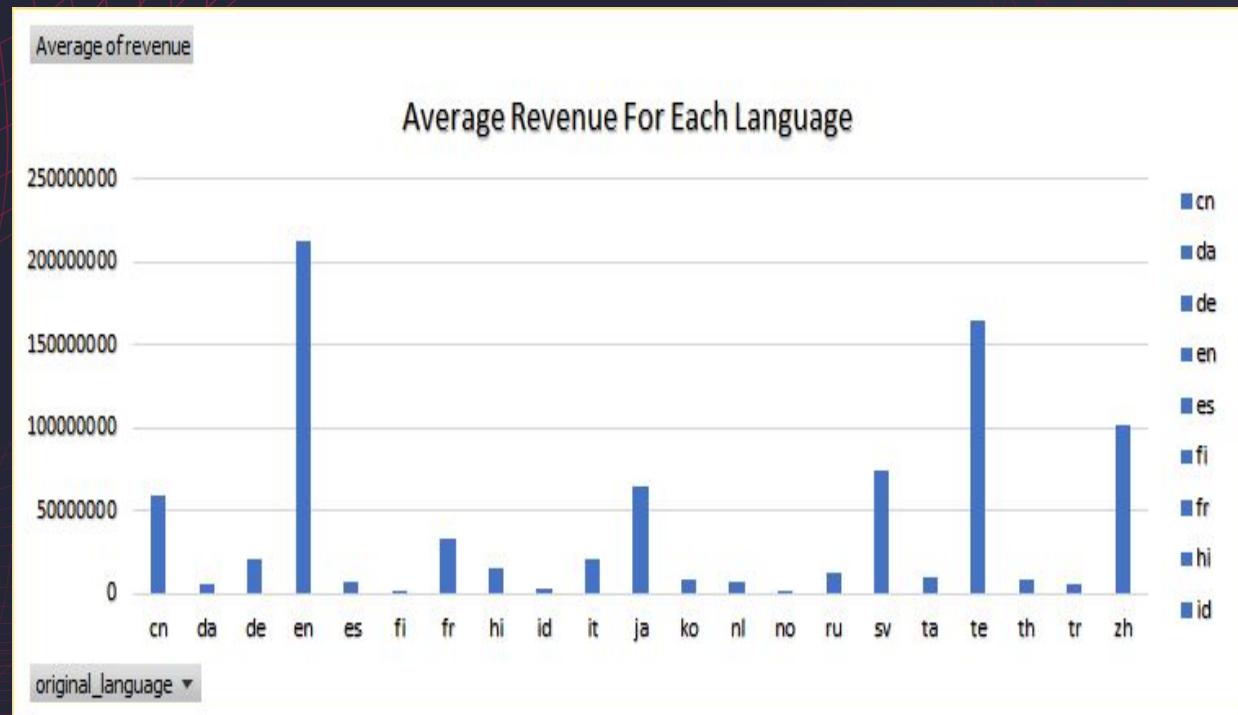
Average Movie Revenues by Original Languages

1.EN (212.3 Millions)

2.TE (165 Millions)

3. Zh (102.2 Millions)

4. JA (65.3 Millions)

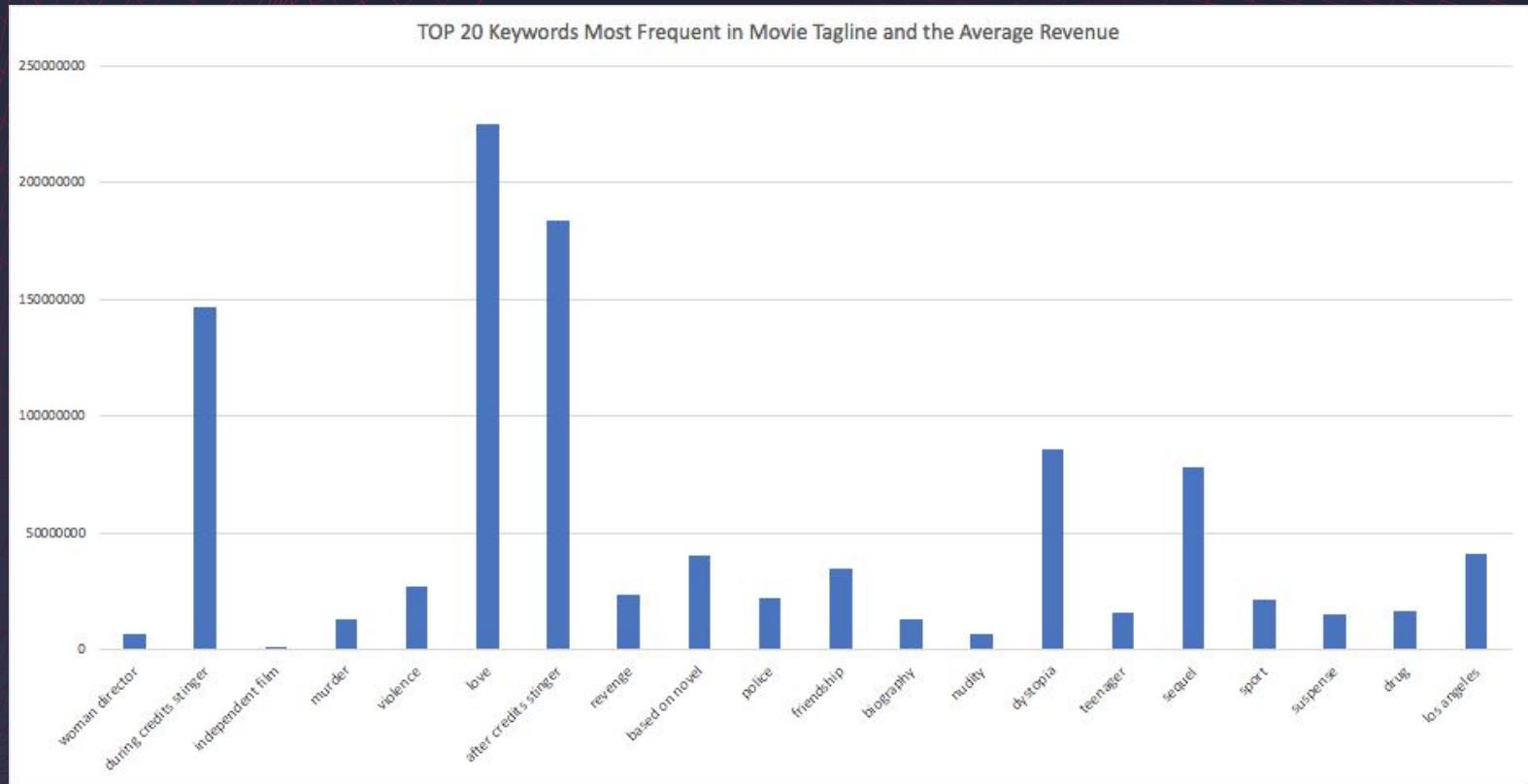


KEYWORDS

How the selection of keywords in description tagline contribute to a movie's revenue



Top 20 Keywords and Their Associated Average Revenue Generated



Excel Regressions

How the selection of genres
contributes to a movie's revenue



03

Model and Final Summary

Data Insight Conclusion and Movie
Investment Strategy Recommendations

MODEL PREDICTION

Model:

- Logistic Regression

Purpose

- Predicting if a movie will be successful (higher than average ROI 404%)

Attributes Include:

- Budget (into 4 bins according to 4 quartiles)
- Top 3 Genres
- Top 10 Production Companies
- Top 10 Actors/Actresses
- Top 10 Keywords
- Top 10 Directors

I. Confusion Matrix and Accuracy

Testing Confusion Matrix

Count 200	Predicted Yes	Predicted No
Actual Yes	TP: 135	FN: 1
Actual No	FP: 4	TN :60
Accuracy:		97.50%

Data-Based Finding Summary & Recommendations

- Most High Revenue movies are released in December and the Summer Season
- On average, movies with a keyword “love” have the highest generated revenue
- English language has the highest average revenue and most of our analyses are based on movies with English language
- The two top performing actors include Ian McKellen and Emma Watson
- Prediction based on chosen variable inputs to accurately measure the success of the movie



Recommendations

- Invest heavily in marketing as popularity of a movie is highly correlated to the revenue of a movie
- Release the movie during summer or December
- Use the set of directors and actors/actresses that yield the highest ROI on average
- The logistic regression model will allow a more customizable recommendation (testing out which combinations of attributes of a movie would possibly yield the highest ROI for different budget bins)

THANKS!

Does anyone have any questions?